

Paris, IAP - 16 juin 2006



Workflow

Réflexion et développements menés au CDS
JLOW, AIDA, cas d'utilisation,...

J.-J. Claudon, C. Pestel, A. Schaaff



JLOW pour la représentation de workflow

■ Développement de bibliothèques « lite »

■ Java Libraries fOr Workflow

■ Côté client

- Basé sur JGraph (<http://www.jgraph.com/>)
- Sauvegarde de la représentation dans le format XML de JGraph

➤ Possibilité d'extension du format

■ Côté serveur

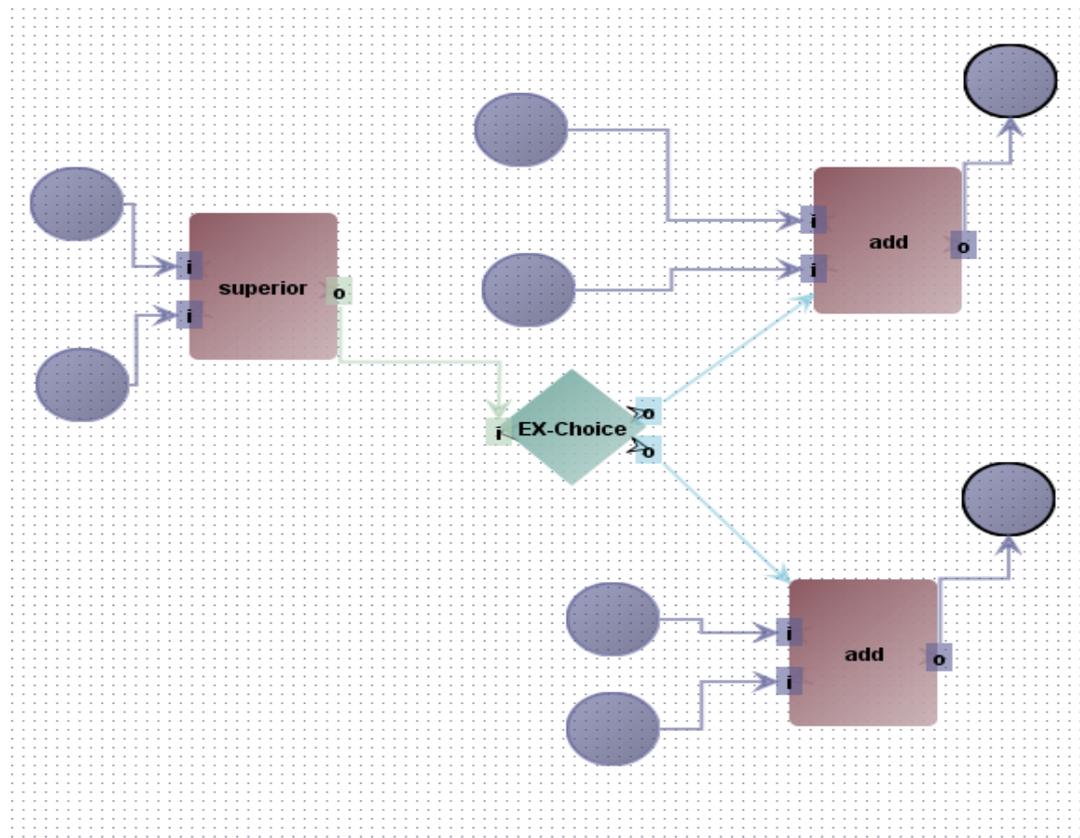
- Très léger pour l'instant mais suffisant pour une intégration dans une application comme AIDA
- Il est envisageable d'utiliser un autre moteur
- ...

■ JLOW est développé dans le cadre du projet VOTech

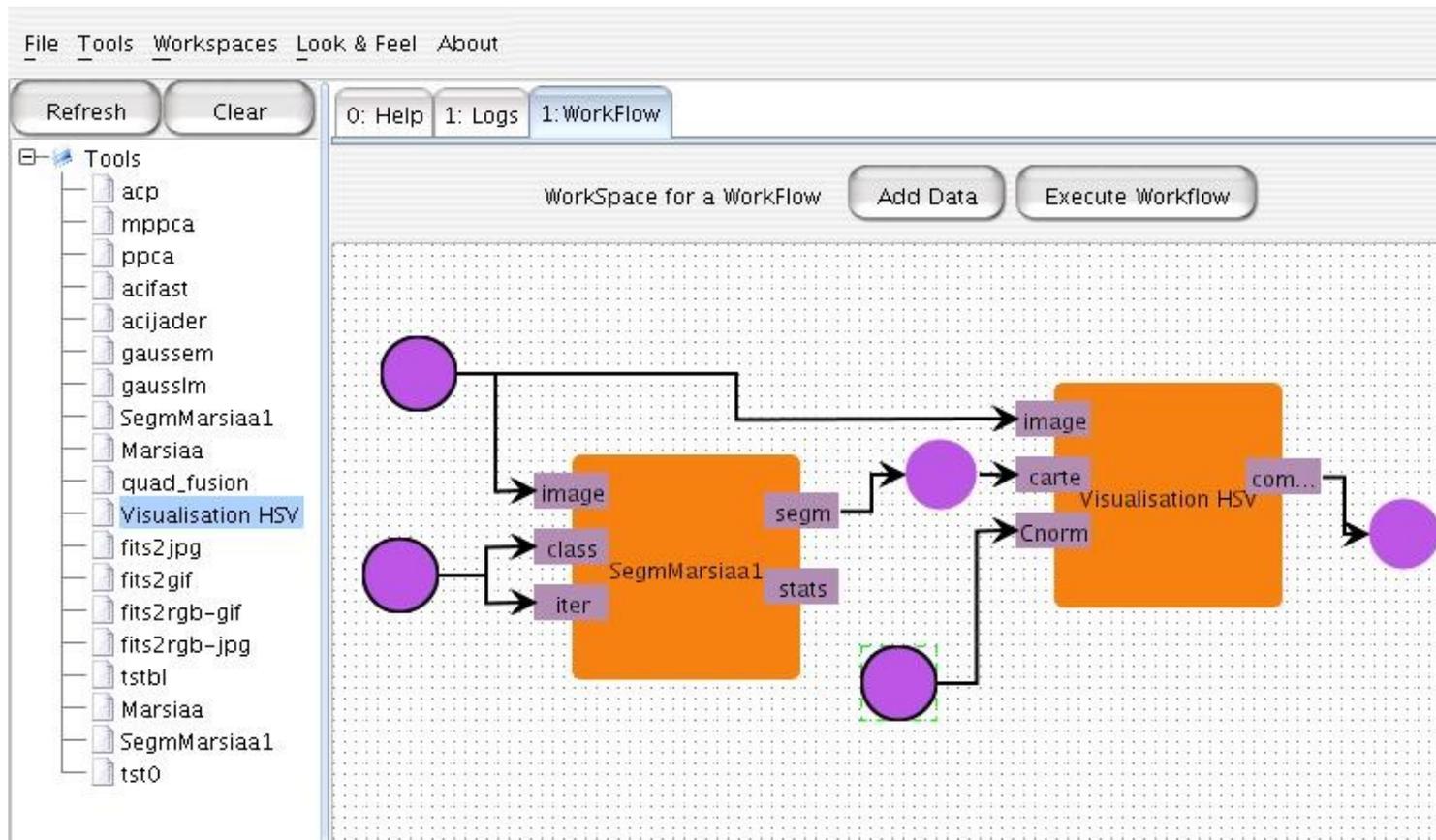


Cas d'utilisation trivial

- JLOW n'est pas une application mais un ensemble de bibliothèques à implémenter suivant ses propres besoins



Cas d'utilisation dans AIDA



AIDA est développé dans le cadre du projet MDA (3,5 ans, fin en décembre 2006) de l'ACI MD

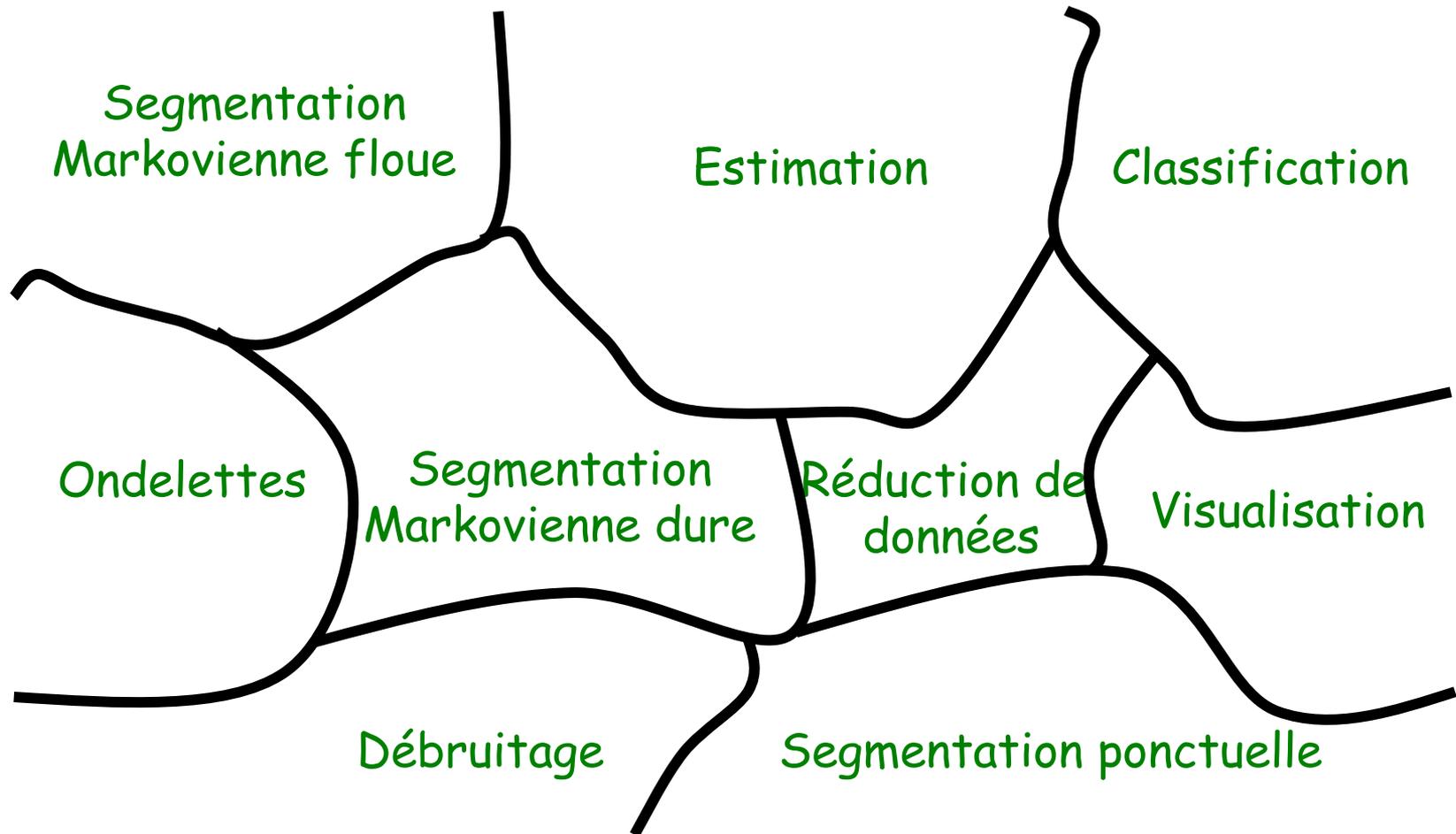


Origine d'AIDA

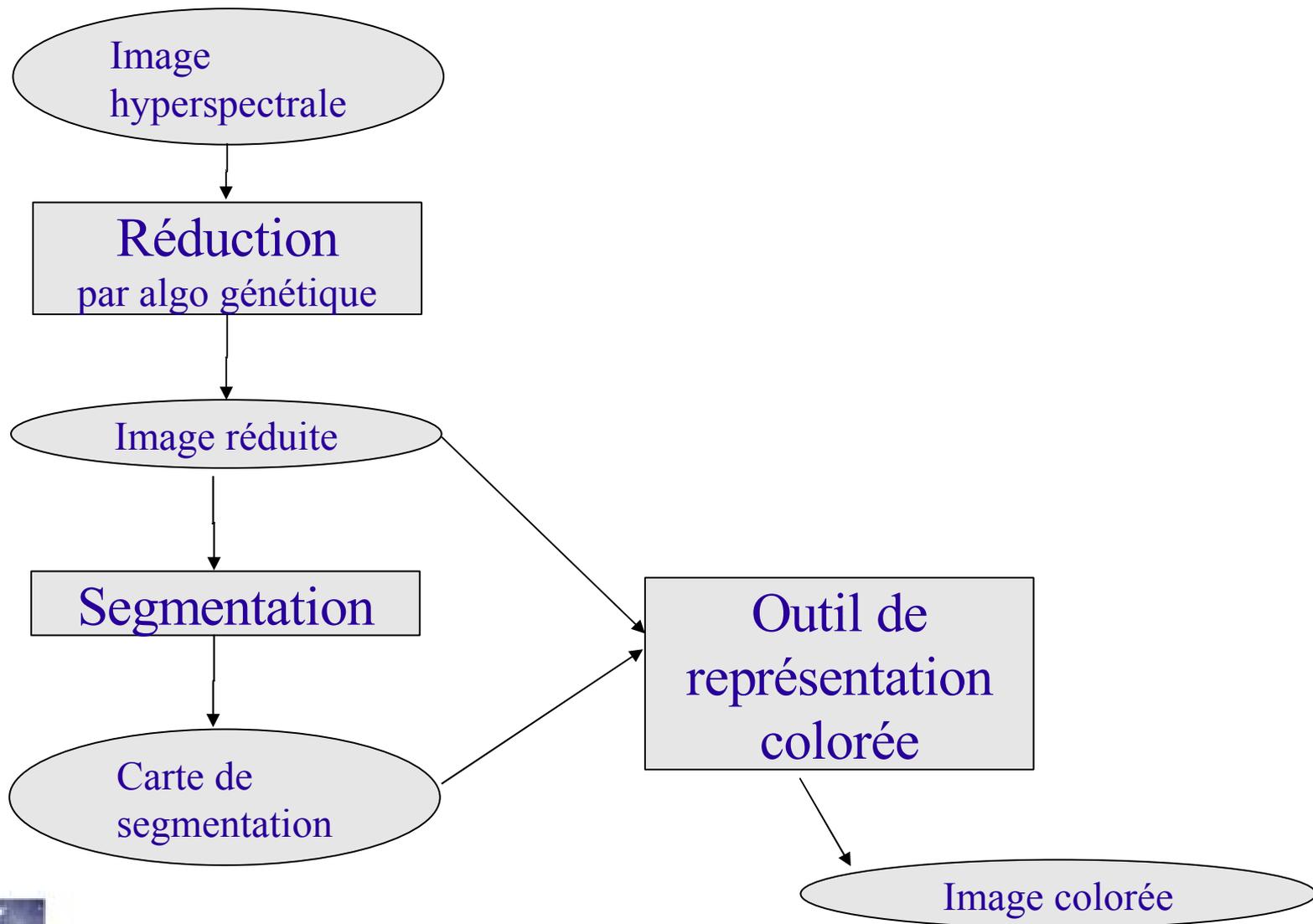
- **Utilisation du traitement d'image en Astronomie**
- **Outils existants**
 - **Différentes techniques de traitement implémentées essentiellement en Matlab (contraintes : exécution locale et licence propriétaire) et dans un contexte de recherche**
 - **Utilisation au cas par cas :**
 - par un public restreint (nouveaux algorithmes de traitement et cas d'étude astronomique)
 - en optimisant principalement les algorithmes mais les temps d'exécution restent élevés



Des programmes de traitement d'images



Un cas d'utilisation



Objectifs initiaux

- **Mise à disposition uniformisée des outils**
 - Interface utilisateur
 - Ouverture vers d'autres outils (Aladin, ...) par des méthodes d'accès génériques
 - Intégration facilitée de nouveaux outils

- **Amélioration à divers niveaux**
 - Performances
 - Enchaînement de tâches
 - Diminution de la complexité et de l'hétérogénéité

- **Utilisation élargie à un plus vaste public**



Mise à disposition uniformisée d'outils

■ Au niveau de l'utilisateur

- Masquer les langages utilisés
- Proposer des traitements sous forme de « boîte noire » avec des entrées et des sorties
 - Référentiel des outils disponibles, sélection d'un outil, génération d'une interface utilisateur spécifique, ...

■ Au niveau de l'intégration des outils

- Apporter à des outils écrits dans divers langages (MATLAB, C, ..) une interface standard (cgi-bin, Web Service)
 - Utilisation d'un fichier « descripteur »
 - Intégration dans un serveur Web
 - induit l'ouverture vers d'autres outils (exemple : Aladin)
 - permet la mise en œuvre d'enchaînements de tâches (Workflow)



Améliorations ...

■ Première piste

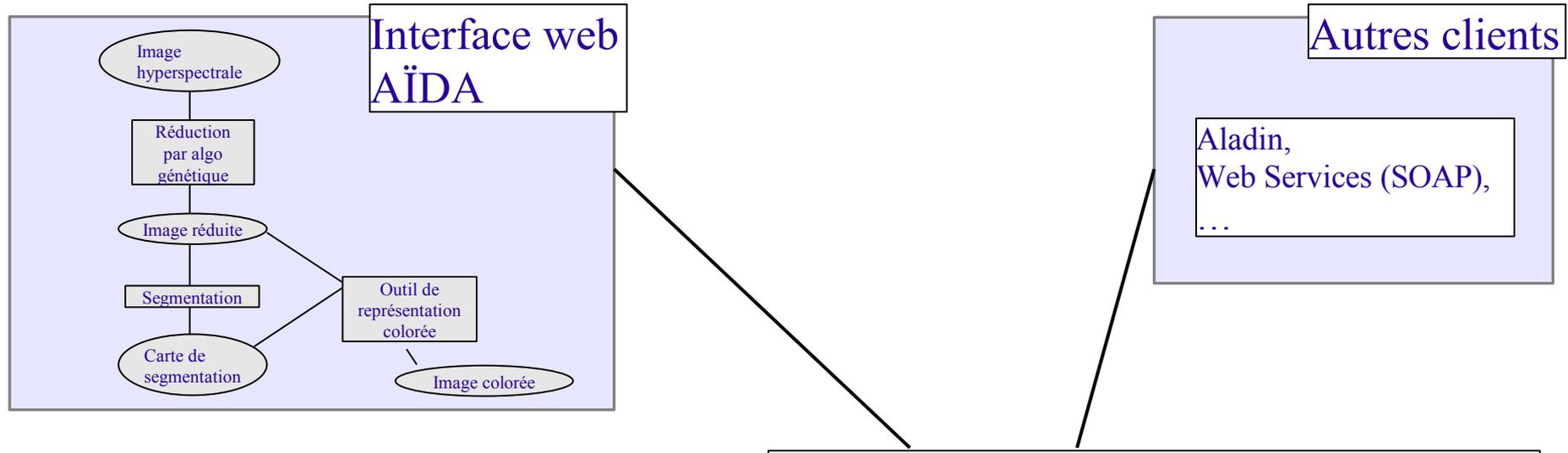
- **Modification des sources C générés à partir d'utilitaires MATLAB**
 - Parallélisation de certaines parties des sources abandonné pour diverses raisons : travail à reproduire après modification des programmes écrits en MATLAB, sources masqués dans les dernières versions de MATLAB, ...

■ Solution actuelle

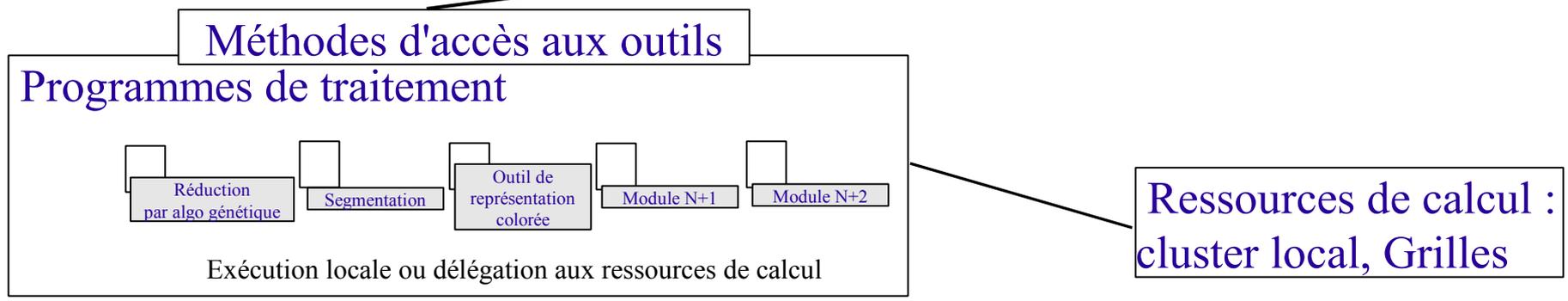
- **Prise en compte en amont (dans MATLAB) du découpage (dans la mesure du possible) des traitements**
- **Intégration de la notion de « Workflow » dès le développement initial**
 - Par rapport aux traitements et par rapport aux données (découpage d'une image en n parties)
- **Exploration des techniques de Grilles**



AIDA : Astronomical Image processing Distribution Architecture



=> intégrateur d'outils



Ressources de calcul (1) : le cluster local

■ Nouvelle configuration

- 8 biprocesseurs Opteron 64 bits
- 2 biprocesseurs Opteron 64 bits avec 3,2 To de disques
- Travail d'intégration en cours, en partie dans le cadre d'un stage MDA de 6 mois (Cyril Pestel, Master2)



Ressources de calcul (2) : exploration des techniques de Grilles

- **Un séminaire et une réunion de travail avec Johan Montagnat et Tristan Glatard (ESSI Sophia)**
 - Premiers tests d'accès à une grille (Grid5000)
 - Accès aux grilles au travers de l'outil « MOTEUR » à explorer plus avant
- **Utilisation des grilles à bon escient**
 - Intéressante lorsque le traitement est lourd et qu'il peut être découpé
 - Cas asynchrone
 - Inutilisable dans le cas de traitements interactifs (problème de maîtrise de l'ordonnancement, ...)
- **Autre piste (hors grille) : Kerrighed (Renaud Lottiaux, IRISA) (intéressant mais ne sera pas utilisable en production avant environ 18 mois)**



■ Un cluster local est indispensable

- Pour les traitements pas trop coûteux en ressources de calcul
- Pour les traitements coûteux mais non parallélisables
- Pour garantir de bonnes performances lors de l'utilisation interactive d'outils

■ Le recours aux grilles peut se justifier dans certains cas :

- Traitements parallélisables et très coûteux en ressources de calcul
 - Risque de dégradation des performances du cluster au détriment des autres tâches
 - Adapté aux grilles qui permettent un changement d'échelle au niveau du nombre de nœuds utilisables



Résultats

- Amélioration notable de l'ergonomie et des performances
- Aide à l'élaboration de « workflow »
- Proposition d'outils pointus à des astronomes non spécialistes en traitement d'image
- ...



Cas d'utilisation au niveau des services

■ Groupe de travail interne

■ Réflexion à propos de l'utilisation de workflow

■ Exemple pour Aladin :

- on dispose d'un catalogue de sources radio
- on souhaite rechercher des contreparties dans le catalogue SDSS
- on "cross-match" les 2 catalogues
- pour chaque ligne du fichier VOTable résultat
 - on charge, centré sur la position du catalogue SDSS, un cutout SDSS et un cutout First
 - on crée une image colorée en combinant les 2 cutouts
 - l'image est ensuite analysée par un astronome qui, d'après ce qu'il voit, va décider de poursuivre une branche ou une autre du workflow.

■ Travail en cours pour VizieR, Simbad, ...,



Perspectives

■ Architecture plus ouverte :

■ Possibilité d'ajout de traitements délocalisés

- Vers une notion de nœud AIDA
- Création de « workflow » comportant des traitements répartis sur plusieurs nœuds (installation d'AIDA à l'Observatoire de Nice)
- Etude de nouveaux cas d'utilisation de traitements délocalisés

■ Travail prévisionnel

- UWS, VOStore, ...

■ Amélioration de la robustesse du prototype



AïDA

■ Démo

